

HOW PIGEONS DISCRIMINATE THE RELATIVE FREQUENCY OF EVENTS

RICHARD KEEN AND ARMANDO MACHADO

INDIANA UNIVERSITY

This study examined how pigeons discriminate the relative frequencies of events when the events occur serially. In a discrete-trials procedure, 6 pigeons were shown one light nf times and then another nl times. Next, they received food for choosing the light that had occurred the least number of times during the sample. At issue were (a) how the discrimination was related to two variables, the difference between the frequencies of the two lights, $D = nf - nl$, and the total number of lights in the sample, $T = nf + nl$; and (b) whether a simple mathematical model of the discrimination process could account for the data. In contrast with models that assume that pigeons count the stimulus lights, engage in mental arithmetic on numerons, or remember the number of stimuli, the present model assumed only that the influence of a sample stimulus on choice increases linearly when the stimulus is presented, but decays exponentially when the stimulus is absent. The results showed that, overall, the pigeons discriminated the relative frequencies well. Their accuracy always increased with the absolute value of the difference D and, for $D > 0$, it decreased with T . Performance also showed clear recency, primacy, and contextual effects. The model accounted well for the major trends in the data.

Key words: relative frequency discrimination, recency and primacy effects, context, mathematical model, key peck, pigeon

Time, space, and number are primitive aspects of the world surrounding most vertebrates. Not surprisingly, then, pigeons, rats, chimpanzees, and humans, among other species, are sensitive to temporal and spatial relations among events and to their numerosity (for reviews, see e.g., Boysen & Capaldi, 1993; Davis & Perusse, 1988; Gallistel, 1990; and see also Boysen & Bernston, 1989; Pepperberg, 1987; Roberts, 1995). Of these three basic aspects, number is probably the least investigated, and hence the least understood, by psychologists. The present study fills some of this gap by reporting an experiment on numerosity, or relative frequency, discrimination in pigeons, and by suggesting a simple quantitative model to interpret some empirical findings on this matter.

Relative frequency discrimination tasks may be divided into two categories according to whether the relevant stimulus events are presented simultaneously or successively. Honig and Matheson's (1995) study illustrates the first category: Pigeons were pre-

sented with an array of 36 blue and red squares and received food for choosing a left key when there were more red squares and a right key when there were more blue squares. The second category is illustrated by Alsop and Honig's (1991) and Machado and Cevik's (1997) studies: Pigeons were presented lights of different colors, one at a time, with reinforcement contingent on choosing the most frequent (Alsop & Honig, 1991) or the least frequent (Machado & Cevik, 1997) sample stimulus. These two studies showed effective frequency discrimination when two or three colors are used as sample stimuli, when the samples range from 5 to 20 stimuli, and when the order of the stimuli within the sample is varied. They also showed that responding generalized to new, untrained samples.

The present study also dealt with the discrimination of relative frequency in a sequential task; it had two major goals. The first was to determine how the discrimination is functionally related to the *difference* (D) between the number of elements in each set of the sample and to the *total* (T) number of elements in the sample. If a sample comprises, say, 16 red and green lights, how does performance vary with the difference between the number of red and green lights (i.e., same T , different D)? Conversely, for a constant difference between the number of red

This research was supported by a FIRST award grant from the National Institute of Mental Health to the second author. We thank Francisco Silva for many helpful comments on earlier versions of the paper.

Address correspondence to Armando Machado, Department of Psychology, Indiana University, Bloomington, Indiana 47405 (E-mail: amachado@indiana.edu).

and green lights, say four, how does performance vary with the total sample size (i.e., same D , different T)? Both Alsop and Honig (1991) and Machado and Cevik (1997) found that accuracy typically increased with D , and some of their findings seemed to indicate that accuracy decreases with T . Together they suggest that accuracy may increase with the ratio D/T , a sort of Weber fraction in the domain of numerosity. However, neither study varied systematically, independently, or over a large range the two variables, D and T . Hence, in the experiment reported below, numerosity discrimination was studied with samples that ranged from $T = 4$ to $T = 28$ stimuli and with differences that ranged from $D = -14$ to $D = 0$ to $D = 14$.

The second goal of the present study was to develop a quantitative model of numerosity discrimination in sequential tasks and to test the model against the data. The model was inspired by the following set of empirical results.

1. As mentioned above, the accuracy of relative frequency discrimination increases with the difference between the number of the distinct stimuli that comprise the sample, but it seems to decrease with the total number of stimuli in the sample.

2. Relative frequency discrimination shows a recency effect, that is, stimuli that occur later in the sample exert greater control over the choice response than do stimuli that occur earlier (Alsop & Honig, 1991; Machado & Cevik, 1997; Roberts & Grant, 1974). Metaphorically speaking, the numerosity of a stimulus seems to inflate when that stimulus is recent. Recency effects also occur in many other tasks, and they may simply express the fact that stimulus control decays with time (Hitch, 1983).

3. Occasionally, relative frequency discrimination also shows a primacy effect; that is, the first stimulus of the sample exerts greater control over the choice response than do subsequent stimuli (e.g., Machado & Cevik, 1997). The significance of the primacy effect remains uncertain, however. First, it is less reliable than the recency effect. Second, it is unclear what process could generate it. Some authors, for example, have suggested a sort of perceptual contrast between the intertrial interval and the beginning of the trial (e.g., Donahoe & Palmer, 1994); others have sug-

gested some form of proactive interference wherein the effect of later stimuli is disrupted by previous ones (e.g., Grant & Roberts, 1973); still others have argued that most primacy effects reported in the literature are due to methodological and statistical artifacts (Gaffan, 1992; but see Wright, 1994; Wright, Santiago, Sands, Kendrick, & Cook, 1985).

Before we describe how this set of results was incorporated into a model, we identify the general attributes of the discrimination task covered by the model. Assume that during a sampling period a pigeon is presented with two stimuli, a red (R) or a green (G) keylight. The stimuli occur in two blocks, with all red lights either preceding or following the green lights (e.g., Alsop & Honig, 1991; Grant & Roberts, 1973). The analysis and interpretation of the results are greatly facilitated when the stimuli are not intermixed because the number of distinct samples is reduced to two. Thus, with five red and seven green lights there are only two samples, RRRRRGGGGGG and GGGGGGRRRRR, whereas if the stimuli are presented randomly the number of samples is 396.

We refer to the frequency of the first and second stimuli as nf and nl , respectively. The total sample size T will then equal $nf + nl$ (12 in the examples above), and the difference between the two frequencies D will equal $nf - nl$ (-2 or $+2$ in the examples). Unless otherwise stated, D always refers to the *signed* difference between the frequencies of the first and last stimuli. After the sample, the pigeon is given a choice between the two keylights, with reinforcement contingent on choosing the keylight that was presented the fewest number of times during the sample (R in the example).

To model the discrimination process we assume that the two stimuli compete for control over the choice response. The degree of control exerted by each stimulus—what we call its control function—increases by a constant amount with each presentation of the stimulus and decreases exponentially when the stimulus is absent. At the moment of choice, the animal chooses one or the other stimulus according to the current values of the two control functions. Specifically, let the degree of control exerted by the first and second stimuli of the sample equal S_F and S_L , respectively. At the beginning of the trial, $S_F =$

$S_L = 0$. Then, after nf instances of the first stimulus,

$$S_F = \beta_1 \times nf, \quad (1)$$

where $\beta_1 > 0$ represents the increment in the control function of the first stimulus with each presentation of that stimulus. We assume a linear function because most experiments on number discrimination seem to support it (e.g., Mechner, 1958; Meck & Church, 1983; see Gallistel, 1990, for a review).

Thereafter, while the second stimulus is presented, S_F decreases exponentially at some rate α . The decay of S_F can be due to two kinds of factors: retroactive interference and the passage of time. If each occurrence of the second stimulus is conceived as a brief "impulse" that reduces the value of S_F by a proportional amount, and if time decay follows an exponential law as many investigators have suggested (Killeen, 1994; Roberts & Grant, 1974; Wixted, 1990), then the combined effect of retroactive interference and time decay will also be exponential (see the Appendix). Hence, after nl instances of the second stimulus,

$$S_F = (\beta_1 \times nf) \exp(-\alpha \times nl). \quad (2)$$

The reason to use nl as the amount of time during which S_F decays is that when the sample stimuli are presented at regular (or approximately regular) intervals, the total time of decay is proportional to nl , and the constant of proportionality can be absorbed by α . Henceforth, *decay* will refer to the decrease in S_F during the presentation of the second stimulus regardless of whether the decrease is due to the passage of time, retroactive interference, or, most likely, to both factors.

With each occurrence of the second stimulus the value of S_L —that is, the degree of control exerted by the second stimulus—increases by a constant amount. Hence, at the end of the sample,

$$S_L = \beta_2 \times nl, \quad (3)$$

where $\beta_2 > 0$.

In principle, the proportionality constants β_1 and β_2 can be related in one of three ways, $\beta_1 = \beta_2$, $\beta_1 > \beta_2$, or $\beta_1 < \beta_2$. If $\beta_1 = \beta_2$, all stimulus occurrences have the same effect. If $\beta_1 > \beta_2$, the first stimulus reduces the effect of each occurrence of the second stimulus, a

case of proactive interference. In this case the processes of interference and decay counteract each other because the former reduces the control of the last stimulus but the latter increases it. If $\beta_1 < \beta_2$, the second stimulus reduces the effect of each presentation of the first stimulus. But it is hard to conceive how a stimulus that has not yet occurred can change the effect of the current stimulus. Alternatively, one could say that the first stimulus somehow increases the effect of each occurrence of the second stimulus, but then the stimuli would not be in competition, as we assumed before. Hence, on the basis of the model's interpretation, we will assume that only the first two cases are possible, that is, $\beta_1 \geq \beta_2$.

During the choice period the animal chooses the last stimulus of the sample as the least frequent stimulus with probability

$$p(\text{last}) = \frac{S_F}{S_F + S_L}. \quad (4)$$

When S_F is larger than S_L , $p(\text{last}) > .5$ and the animal is more likely to choose the second stimulus; when S_L is larger than S_F , $p(\text{last}) < .5$ and the animal is more likely to choose the first stimulus. The form of Equation 4 is determined by the fact that because the experimenter reinforces choices of the least frequent stimulus, the stimuli exert avoidance control over the choice response. By avoidance control we mean that each occurrence of a stimulus reduces the animal's propensity to choose that stimulus during the choice period. Hence, all else being equal, the more a stimulus occurs, the less likely the animal is to choose it as the least frequent one.¹

Figure 1 illustrates how the model applies to a particular sample. During the presentation of the first stimulus, five red keylights, S_F increases linearly at rate β_1 . Then, during the presentation of the second stimulus, seven green keylights, S_F decays exponentially at rate α , whereas S_L increases linearly at rate $\beta_2 < \beta_1$. Meanwhile the probability of choosing the last stimulus as the least frequent one (see bottom panel) decreases in a logistic-like

¹ If the experimenter reinforced the choice of the most frequent stimulus, $p(\text{last})$ would equal $S_L / (S_L + S_F)$. See the analysis of Alsop and Honig's (1991) study in the Discussion.

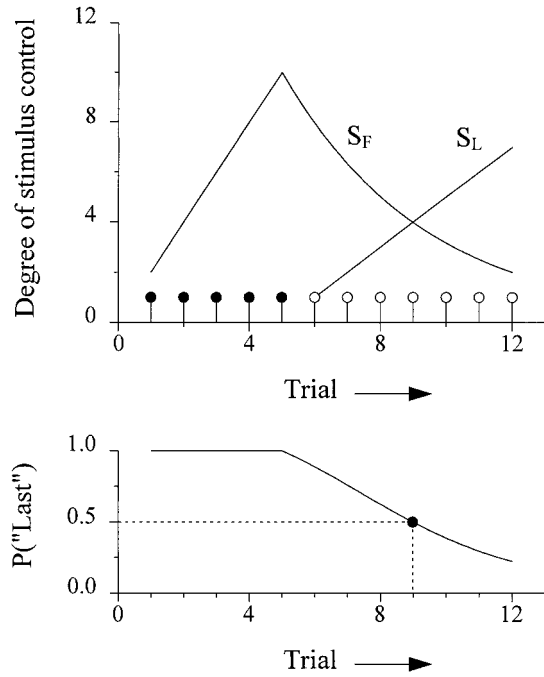


Fig. 1. Model of relative frequency discrimination. Top: Each occurrence of the first stimulus (filled circles) increases S_F by a constant amount, β_1 ; during the second stimulus (open circles), S_F decays exponentially at a rate α . Each occurrence of the second stimulus increases S_L by a constant amount, β_2 . Bottom: The probability of choosing the last stimulus is given by the ratio $S_F/(S_F + S_L)$.

fashion. In this particular example, by the end of the sample the bird would be more likely to choose the correct red keylight.

Given the ratio rule (Equation 4), the model depends effectively on two parameters only, the decay parameter α and the ratio of the proportionality constants β_2 and β_1 :

$$p(\text{last}) = \frac{(\beta_1 \times nf) \exp(-\alpha \times nl)}{(\beta_1 \times nf) \exp(-\alpha \times nl) + \beta_2 \times nl}$$

$$= \frac{1}{1 + \gamma \frac{nl}{nf} \exp(\alpha \times nl)}, \quad (5)$$

where $\gamma = \beta_2/\beta_1$. Because we have assumed that $\beta_1 \geq \beta_2$, γ is expected to be between 0 and 1. The amount by which γ deviates from 1 and approaches 0 can be taken as a measure of the degree of proactive interference. In other words, lower values of γ mean higher degrees of proactive interference. Henceforth, we call γ the interference parameter.

The model incorporates the empirical results listed above. First, as the partial derivatives of $p(\text{last})$ readily show (see the Appendix), accuracy always increases with the absolute difference between the two stimulus frequencies, D , and, for $D > 0$, it decreases with the total number of stimuli, T . Hence, when $D > 0$, accuracy increases with the Weber-like ratio D/T . However, when $D < 0$, the role of T is less clear, for in this case accuracy depends on the absolute values of D , T , and the decay parameter α ; no simple generalization can be made.

Second, to see how the model incorporates the recency and primacy effects, consider the case in which the two stimuli have the same frequency, that is, $nf = nl$. What would a preference for the last stimulus mean? According to the model, it would mean that the first stimulus exerted greater avoidance control than the second. In other words, the first stimulus was considered more frequent than the second even though the frequencies were equal—a primacy effect. Conversely, a preference for the first stimulus would mean that the last stimulus exerted greater control than the first or, in other words, that the last stimulus was considered more frequent than the first—a recency effect. Note that primacy and recency effects are defined not by the stimulus chosen by the animal but by what its choice implies in terms of degree of stimulus control.

To obtain the model's predictions, we first express nf and nl in terms of D and T [i.e., $nf = (D + T)/2$ and $nl = (T - D)/2$], then substitute the resulting expressions in Equation 5, and finally set $D = 0$. The result is the logistic function

$$p(\text{last} | D = 0) = \frac{1}{1 + \gamma \exp(\alpha T/2)}. \quad (6)$$

It follows from Equation 6 that a preference for the last stimulus—a primacy effect—will occur only if $\gamma \exp(\alpha T/2) < 1$. This condition requires both $\gamma < 1$ (i.e., some degree of proactive interference) and a relatively small sample, T ; otherwise, the effect of proactive interference is overridden by the effect of decay. However, regardless of the values of γ and T , the model predicts that the preference for the last stimulus should decrease with the sample size T . Hence, for relatively

large samples the model predicts a preference for the first stimulus—a recency effect.

In the preceding analysis we started with particular samples, namely those in which the two stimuli had the same frequency, and then defined recency and primacy effects in terms of deviations of choice probability from .5 following those samples. A complementary way to define recency and primacy takes the opposite direction: We start with a choice probability of .5—the point of subjective equality—and then examine the samples for which it holds true. Recency and primacy effects are defined by the corresponding values of nf and nl . Thus, if at the point of subjective equality nf is less than nl , then we conclude that the two stimuli influenced the choice response equally even though the first was less frequent than the second—a primacy effect. Conversely, if nf is greater than nl at the point of subjective equality, then we conclude that the second stimulus influenced the choice response as much as the first even though it was less frequent—a recency effect.

The model predicts that at the point of subjective equality nf and nl will be related by the equation $nf = \gamma \times \exp(\alpha \times nl) \times nl$. Hence, a primacy effect (i.e., $nf < nl$) can occur only if $\gamma < \exp(-\alpha \times nl)$. This condition requires $\gamma < 1$ (some degree of proactive interference), and it is more likely to hold as nl becomes smaller. Conversely, a recency effect (i.e., $nf > nl$) requires $\gamma > \exp(-\alpha \times nl)$, which is more likely to hold as nl becomes larger. We have reached the same conclusion: Primacy is more likely after small samples; recency is more likely after large ones.

In summary, the model predicts that (a) accuracy should always increase with the absolute value of D ; (b) for $D > 0$, accuracy should decrease with T ; and (c) a primacy effect may occur for relatively small samples, but for larger samples a recency effect is likely to predominate.

METHOD

Subjects

Six pigeons (*Columba livia*), maintained at approximately 80% of their free-feeding body weights, participated in the experiment. They were housed in individual home cages with

free access to water and grit. The pigeon colony was always illuminated. The birds had been used in a previous frequency discrimination experiment and therefore required no preliminary training.

Apparatus

Two identical operant chambers (24 cm by 29 cm by 30 cm) were used. The front panel of the chamber was equipped with three response keys, each 2.5 cm in diameter. The keys were located at the vertices of an inverted isosceles triangle, with the top keys 10 cm apart and 8.5 cm from the center key. The center of the triangle was 22.4 cm above the grid floor. The keys could be lit with green (G), red (R), or blue (B) light. In this experiment, the left key was always red, the right key was always green, and the center blue. An opening (6 cm by 7 cm) 4 cm above the floor allowed access to grain. An outer box enclosed the operant chamber, and a ventilation fan provided air circulation and helped to mask extraneous noises. All experimental events were controlled by a computer.

Procedure

Sessions were divided into trials, and each trial comprised the following series of events. After a 20-s blackout, the houselight and the center key were illuminated, the latter with blue light. Two pecks at the center key turned that keylight off and initiated the sampling period. After a 0.4-s interval, one side key was illuminated and remained so until a peck at that key turned its light off. These events, the 0.4-s interval followed by the illumination of the key and the bird's peck, defined one stimulus occurrence. After the first stimulus of the sample was presented for nf times on one side key, the second stimulus of the sample was presented for nl times on the other side key. The left and right side keys were always illuminated with red and green light, respectively. Following the last 0.4-s interval of the sampling period, the center key was again illuminated with blue light, and the choice period started when a minimum of 2 s elapsed and the pigeon pecked the center key twice.

During the choice period, the center keylight was turned off and both side keys were illuminated. If the bird pecked the side key that was illuminated (and pecked) the fewest number of times during the sampling period,

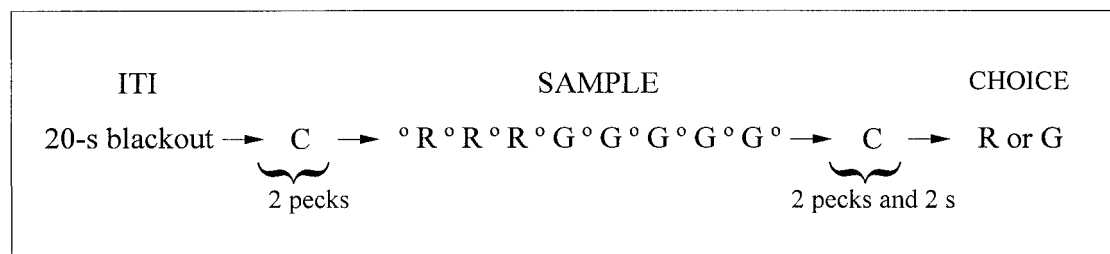


Fig. 2. Trial structure. ITI is the intertrial interval. C refers to the center key, illuminated with blue light; R refers to the left key, illuminated with red light; G refers to the right key, illuminated with green light. The circles refer to the 0.4-s interval.

then it received food for 2.5 s. Afterwards the 20-s blackout started and a new sample was presented. However, if the choice was incorrect, the 20-s blackout started immediately and the same sample was repeated (correction trials method). Figure 2 illustrates the trial structure for the sample three red and five green lights. By using two different lights on two separate keys during the sample period, we attempted to make the discrimination easier to learn; by blocking the stimulus presentations (all red lights first or all green lights first), we reduced the frequency of switching between the side keys and greatly simplified the data analyses; by requiring the bird to peck the center key at the end of the sample, we attempted to eliminate any bias during the choice period (had the choice period followed the sample immediately, the bird would be facing the side key that was illuminated the last when a choice became

available, and that might have biased its responses).

The experiment was divided into four conditions, and the samples used during each one of them are shown in Table 1. Each cell of the table corresponds to one sample, with the first number representing the frequency of the first stimulus, nf , and the second number representing the frequency of the second stimulus, nl . During Condition 1, T ranged from 4 to 16, and D ranged from -14 to $+14$. Condition 1 lasted from 46 to 50 sessions. (Although proportion correct stabilized in less than 12 sessions, a relatively large number of sessions was needed to estimate the choice probabilities reliably.) During Condition 2, which lasted from 47 to 51 sessions, T ranged from 16 to 28, but the range of D remained the same, from -14 to $+14$. Note that the $T = 16$ samples were equal during Conditions 1 and 2 and constituted the same fraction

Table 1

Stimulus frequencies used during the experiment. The first and second numbers of each cell refer to the frequencies of the first (nf) and second (nl) stimuli of the sample, respectively. $D = nf - nl$; $T = nf + nl$.

		Conditions 1, 3, and 4				Condition 2			
						T			
		4	8	12	16		16	20	24
D	-14				1, 15	1, 15	3, 17	5, 19	7, 21
	-10			1, 11	3, 13	3, 13	5, 15	7, 17	
	-6		1, 7	3, 9	5, 11	5, 11	7, 13		
	-2	1, 3	3, 5	5, 7	7, 9	7, 9			
	0 ^a	2, 2 ^a	4, 4 ^a	6, 6 ^a	8, 8 ^a				
	+2	3, 1	5, 3	7, 5	9, 7	9, 7			
	+6		7, 1	9, 3	11, 5	11, 5	13, 7		
	+10			11, 1	13, 3	13, 3	15, 5	17, 7	
	+14				15, 1	15, 1	17, 3	19, 5	21, 7

^a These samples were used only during Condition 4.

(40%) of the total number of samples. However, the $T = 16$ samples were the largest during Condition 1 but the smallest during Condition 2. Hence, if frequency discrimination is influenced by the overall context of a sample, then we should detect such an effect with the $T = 16$ samples. Condition 3 was an exact replication of Condition 1, and it lasted from 46 to 52 sessions. Finally, Condition 4 was identical to Conditions 1 and 3, except that it also contained samples in which the stimulus frequencies were equal (i.e., $nf = nl$, and therefore $D = 0$). When the stimulus frequencies were equal, no choice response was reinforced. Condition 4 lasted for 20 sessions.

Each sample was presented four times per session, twice with the red (left) key as the first and the green (right) key as the second stimulus of the sample, and twice with this order reversed. Moreover, all samples in the filled cells of Table 1 were presented during each session of the corresponding condition. Hence, during Conditions 1, 2, and 3, sessions ended after 80 trials (20 samples \times 4 trials per sample, excluding correction trials) and during Condition 4 they ended after 96 trials (24 samples \times 4 trials per sample).

The data analysis excluded the correction trials and was based on the last 40 sessions of the first three conditions and all 20 sessions of Condition 4. Moreover, because there was very little evidence of bias for a particular key-light color, choice proportions were analyzed as a function of nf and nl only (i.e., regardless of whether nf was red and nl was green or vice versa).

RESULTS

The Roles of D and T in Relative Frequency Discrimination

We report most results in terms of the probability of choosing the last element of the sample, $p(\text{last})$, not in terms of the more familiar probability of a correct response. The reason for choosing this dependent variable is that a single equation, namely Equation 5, summarizes the model's predictions. In contrast, probability correct would require two equations, one for the case $D > 0$ and another for the case $D < 0$. Moreover, probability correct, but not $p(\text{last})$, is undefined when $D = 0$.

Condition 1. Figure 3 shows the data from Condition 1. The probability of choosing the last stimulus is plotted as a function of D . To assess the effect of D , we maintain T constant and read the results along each curve. For every sample size, the probability of choosing the last stimulus increased with D , and because $p(\text{last})$ is equivalent to proportion correct when $D > 0$ but to proportion incorrect when $D < 0$, this result implies that accuracy increased with the absolute value of the difference between the stimulus frequencies.

To assess the effect of T , we maintain D constant and read the results vertically. When $D > 0$, proportion correct decreased with the sample size. This result was highly reliable, in that all birds showed it without a single inversion. However, when $D < 0$, the effect of the sample size was less clear. The results for $D = -2$ showed that proportion correct [i.e., $1 - p(\text{last})$] increased with larger samples, the opposite of what was found for $D > 0$, but the data points at $D = -10$ and $D = -6$ were too close to one another and to 0 to warrant any conclusion.

The curves show the model's fit with the parameter values and the variance accounted for displayed in Table 2. Here and in subsequent conditions, the model was fit to the entire data set of each bird. Although it was possible to fit the model to each sample size separately, we reasoned that the model would be far more convincing if it could fit all data points simultaneously. In addition, by fitting the model to all data points at once, more weight was given to the independent variables D and T and less weight was given to the model's free parameters γ and α . The curves on the bottom panel were obtained by averaging the individual curves, not by fitting the average data anew.

The variance accounted for by the model was always substantial ($\omega^2 = .988$), and no systematic differences between model and data were apparent. The interference and decay parameters had reasonable and consistent values. In particular, the interference parameter was less than 1 for all pigeons, a result that indicates the presence of proactive interference. (The reader is reminded that low values of γ correspond to higher degrees of proactive interference.) The absolute values of γ also show that, according to the model's interpretation, the control function of the sec-

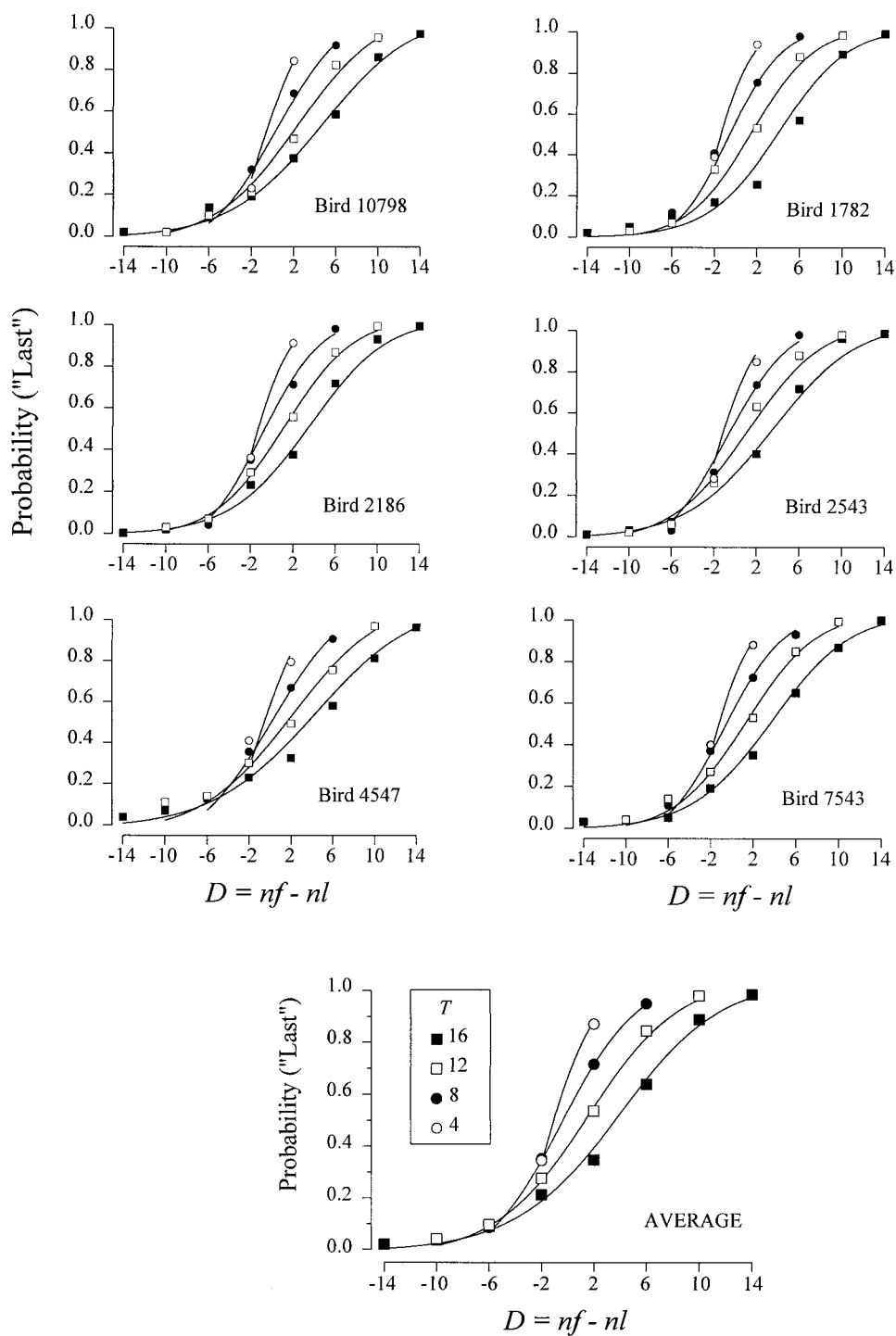


Fig. 3. The symbols show the probability of choosing the last stimulus of the sample as a function of the difference between the stimulus frequencies, D . Each different symbol corresponds to a different total sample size, T . The top panels show the data from individual birds during Condition 1; the bottom panel shows the average across birds. The curves are the model's predictions using the parameters shown in Table 2.

Table 2

Parameter values used to fit the model (α and γ) and variance accounted for (ω^2) in each condition.

Bird	Condi- tion	γ	α	ω^2	Bird	Condi- tion	γ	α	ω^2
10798	1	0.438	0.227	.993	2543	1	0.277	0.264	.987
	2	0.170	0.230	.996		2	0.182	0.189	.985
	3	0.246	0.279	.987		3	0.233	0.272	.986
	4	0.377	0.232	.980		4	0.435	0.154	.979
1782	1	0.183	0.362	.984	4547	1	0.498	0.186	.976
	2	0.037	0.407	.994		2	0.522	0.083	.986
	3	0.183	0.323	.990		3	0.353	0.199	.996
	4	0.170	0.376	.975		4	0.429	0.238	.969
2186	1	0.229	0.301	.996	7543	1	0.246	0.299	.994
	2	0.153	0.223	.991		2	0.188	0.210	.993
	3	0.233	0.286	.989		3	0.190	0.330	.990
	4	0.240	0.254	.965		4	0.439	0.210	.986

ond stimulus increased at a rate of 18% to 50% of the rate of the first stimulus.

Condition 2. Figure 4 shows the results for the second condition. The range of the differences between the two stimuli was the same as in Condition 1, but the sample sizes ranged from 16 to 28. However, the pattern of results was similar: As D increased, the probability of choosing the last stimulus also increased, which means that proportion correct varied directly with the absolute difference between the two frequencies. Concerning the effects of T , the probability of choosing the last stimulus usually decreased with T when D was positive. This trend was observed even though the magnitude of the effect was smaller than in Condition 1 and a few reversals occurred. When $D < 0$, the proportions were too close to one another and to 0 to reveal any effect of T .

The model fitted the data well ($\overline{\omega^2} = .991$) and, as the bottom panel shows, there were no systematic differences between model and data. The interference and decay parameters had reasonable values and, once again, the former was always less than 1 (see Table 2). We postpone a more detailed comparison of the data and the parameter values from the first two conditions until we describe the results from the remaining conditions.

Conditions 3 and 4. Figure 5 shows the results from Condition 3. Performance was very similar to that observed in Condition 1. Thus, $p(\text{last})$ increased with D and, for $D > 0$, it decreased with T (again, no reversals were observed). The effect of T was not as clear

when $D < 0$: The average data points at $D = -2$ showed increased accuracy with sample size, but at other values of D they were too close to one another and to 0 to reveal any effect of T . The model accounted for a high proportion of the variance ($\overline{\omega^2} = .990$), and the interference parameter was always less than 1.

Figure 6 presents the results from Condition 4. In addition to the regular samples, this condition also included samples with $D = 0$. The results were slightly more variable than in Conditions 1 and 3, presumably because only half the number of trials was used to estimate $p(\text{last})$. However, the pattern of results, the overall quality of the model's fit ($\overline{\omega^2} = .976$), and the values of the two parameters remained approximately the same.

In summary, the results from all conditions show that relative frequency discrimination varied as an orderly function of the difference between the frequencies of the two stimuli and by the total number of stimuli in the sample. When the first stimulus of the sample outnumbered the second, the discrimination improves with D and worsens with T . In other words, accuracy varies directly with the Weber-like ratio D/T . When the last stimulus of the sample outnumbered the first, the discrimination also improves with D , but the effect of T is weaker, and for that reason is less clear. In the only case in which the data points did not overlap considerably (i.e., $D = -2$) larger samples tended to enhance accuracy, the opposite of what was found for $D > 0$. Concerning the model, it fitted all data sets well, and

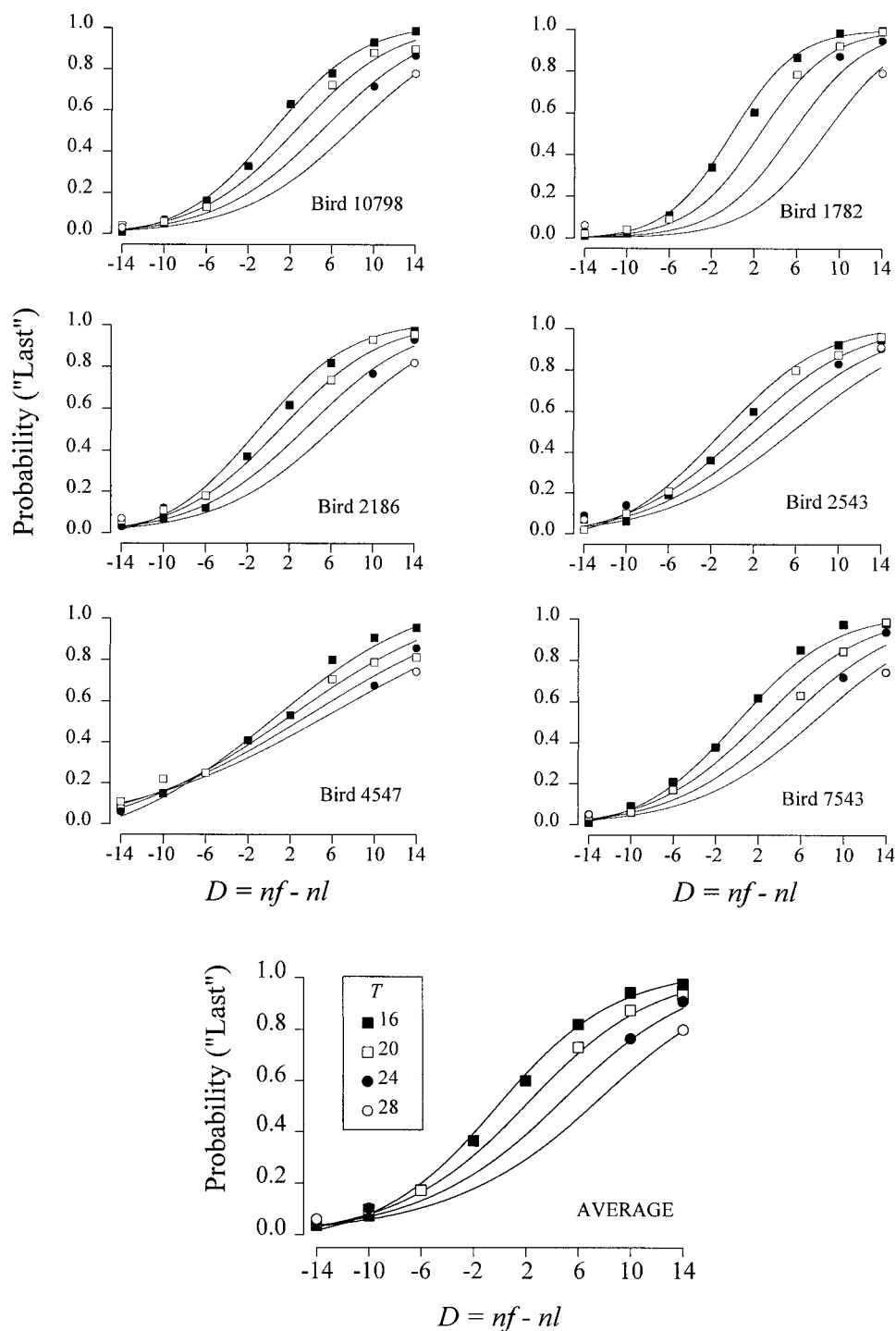


Fig. 4. The symbols show the probability of choosing the last stimulus of the sample as a function of the difference between the stimulus frequencies. Each different symbol corresponds to a different total sample size. The top panels show the data from individual birds during Condition 2; the bottom panel shows the average across birds. The curves are the model's predictions using the parameters shown in Table 2.

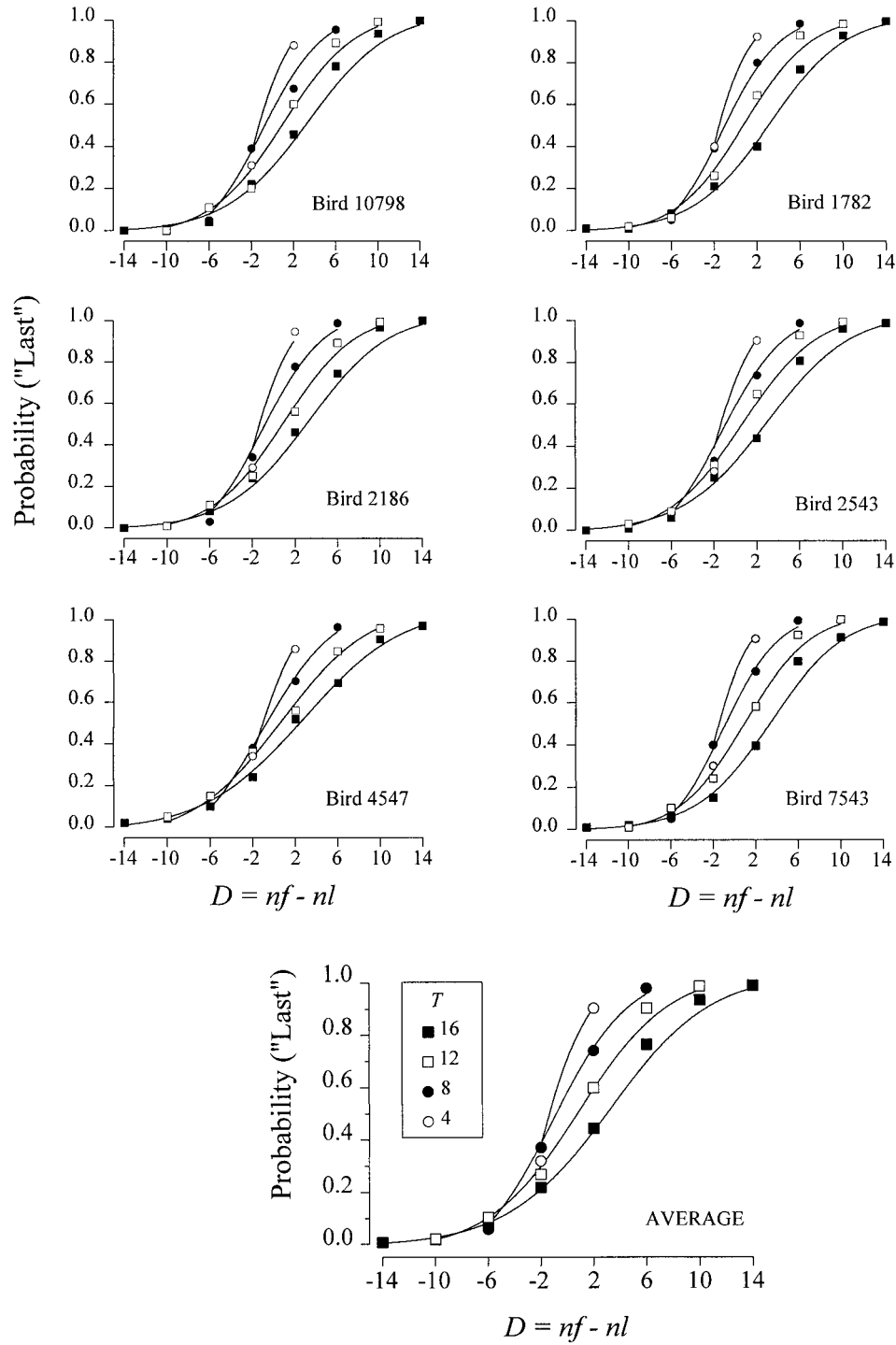


Fig. 5. The symbols show the probability of choosing the last stimulus of the sample as a function of the difference between the stimulus frequencies. Each different symbol corresponds to a different total sample size. The top panels show the data from individual birds during Condition 3; the bottom panel shows the average across birds. The curves are the model's predictions using the parameters shown in Table 2.

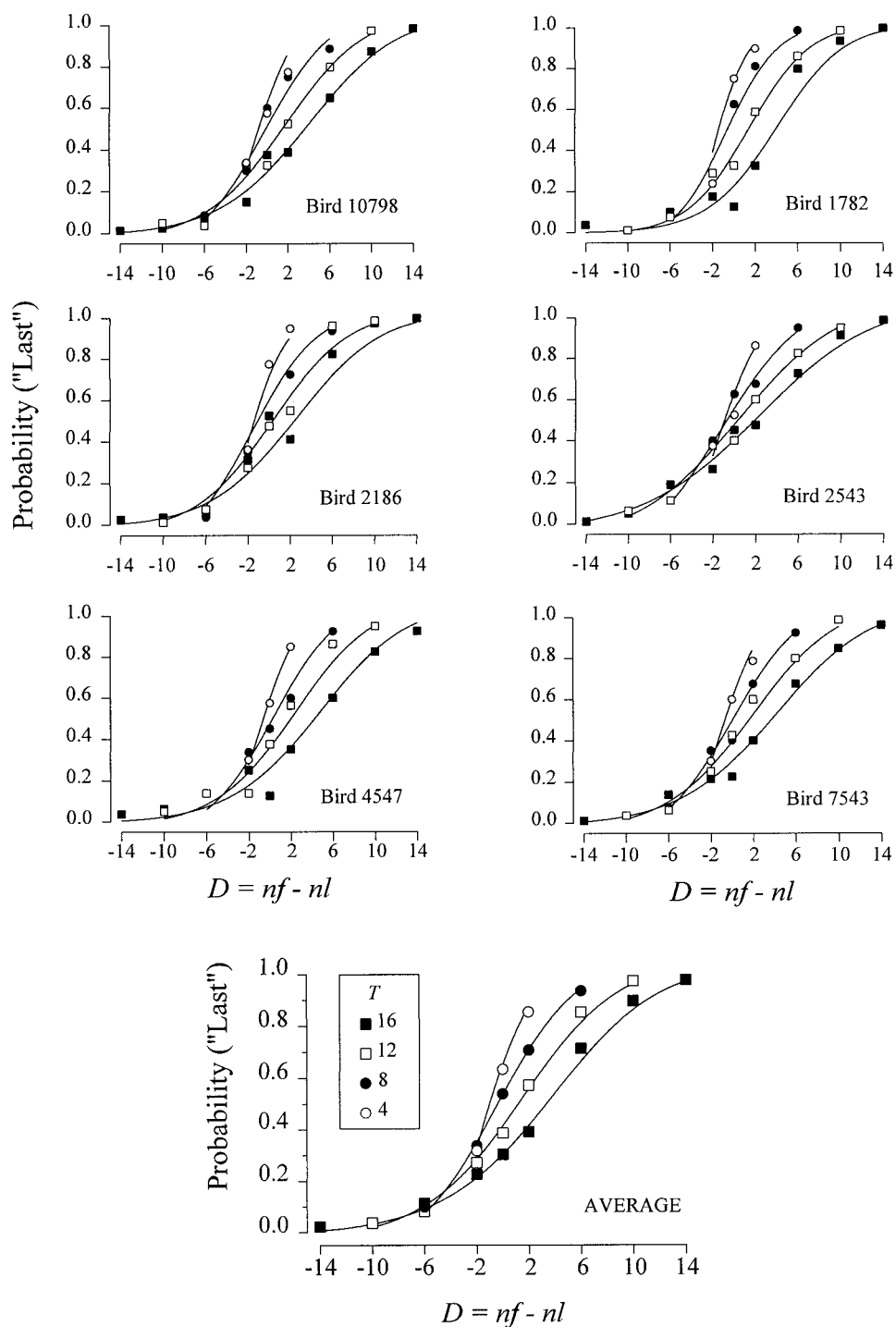


Fig. 6. The symbols show the probability of choosing the last stimulus of the sample as a function of the difference between the stimulus frequencies. Each different symbol corresponds to a different total sample size. The top panels show the data from individual birds during Condition 4; the bottom panel shows the average across birds. The curves are the model's predictions using the parameters shown in Table 2.

its parameters had reasonable and consistent values. In particular, the interference parameter γ was always less than 1, which indicates the presence of proactive interference.

A Context Effect

The fact that all conditions included samples with a total of 16 stimuli allows us to determine whether the general context of a sample influences the discrimination of its component frequencies. Figure 7 plots $p(\text{last})$ for the $T = 16$ samples during Conditions 1 and 2, and the curves show the model's predictions. The results from Conditions 3 and 4 were similar to those from Condition 1 and are not presented.

For all birds, the curve for Condition 2 shifted to the left of the curve for Condition 1. That is, the probability of choosing the last stimulus increased when the average sample size increased from 12 in Condition 1 to 20 in Condition 2. In terms of accuracy, this left shift means that, for $D > 0$, accuracy increased when the average sample size increased, but for $D < 0$, accuracy decreased.

The model provides the following interpretation of this context effect. For 5 pigeons (except 4547), when the average sample size increased in Condition 2, the decay parameter α remained roughly constant, but the interference parameter γ decreased. A smaller value of γ means a higher degree of proactive interference or, in other words, a greater disruptive effect of the first stimulus on the control exerted by the last stimulus. The net effect was a reduced value of S_L at the moment of choice and, consequently, a greater tendency to choose the last stimulus as the least frequent one (see Equation 4).

For Bird 4547, γ remained approximately constant but α decreased during Condition 2. In other words, the degree of proactive interference did not change but the control function of the first stimulus decayed more slowly. The net effect was the same, though: a relatively greater control of the choice response by the first stimulus of the sample. In summary, according to the model the context effect occurred because in Condition 2 the first stimulus of the sample exerted greater control over the choice response. We will return to the context effect in the Discussion.

Recency and Primacy

Primacy and recency effects were measured by two complementary methods, the first based on the points of subjective equality and the second based on the choice probabilities following the samples in which $D = 0$. We describe the results in this order.

The points of subjective equality were estimated as follows. Imagine a horizontal line at $p(\text{last}) = .5$ in each panel of Figures 3 to 6. The point at which the line intersects a curve yields an estimated value of D for a specific value of T . Figure 8 shows for all experimental conditions the average values of this D value as a function of T . Consider the case $T = 4$. The corresponding value of D was approximately -1.1 in Conditions 1, 3, and 4. This value means that the first stimulus had to occur 1.1 fewer times than the last stimulus in order to be considered equally frequent (i.e., $nf - nl = -1.1$). Hence, negative values of D define the primacy effect. By the same logic, positive values of D define the recency effect.

The results indicate that in Conditions 1, 3, and 4, a primacy effect was visible with the smallest samples, whereas a strong recency effect was visible with the largest samples. In Condition 2, however, only recency effects were observed. As the model predicted, the recency effect always increased with the sample size. The discontinuity between the curve from Condition 2 and the curves from the other conditions is another expression of the context effect mentioned before. Samples with 16 stimuli yielded a strong recency effect during Conditions 1, 3, and 4, but no recency or primacy effects during Condition 2.

The second method of measuring recency and primacy effects was based on the deviations of $p(\text{last})$ from .5 following samples with $D = 0$. Because these samples were included only in Condition 4, for the remaining conditions the value of $p(\text{last})$ at $D = 0$ was estimated from the fitted curves. In this case, we imagine a vertical line at $D = 0$ and then determine the point at which the line crosses each curve. We start with the data from Condition 4. The top six panels of Figure 9 reproduce the results from individual birds as well as the model's corresponding predictions. The latter were not a new fit, but resulted from the fit to the entire data set of Condition 4 (see Figure 6). For all birds, the

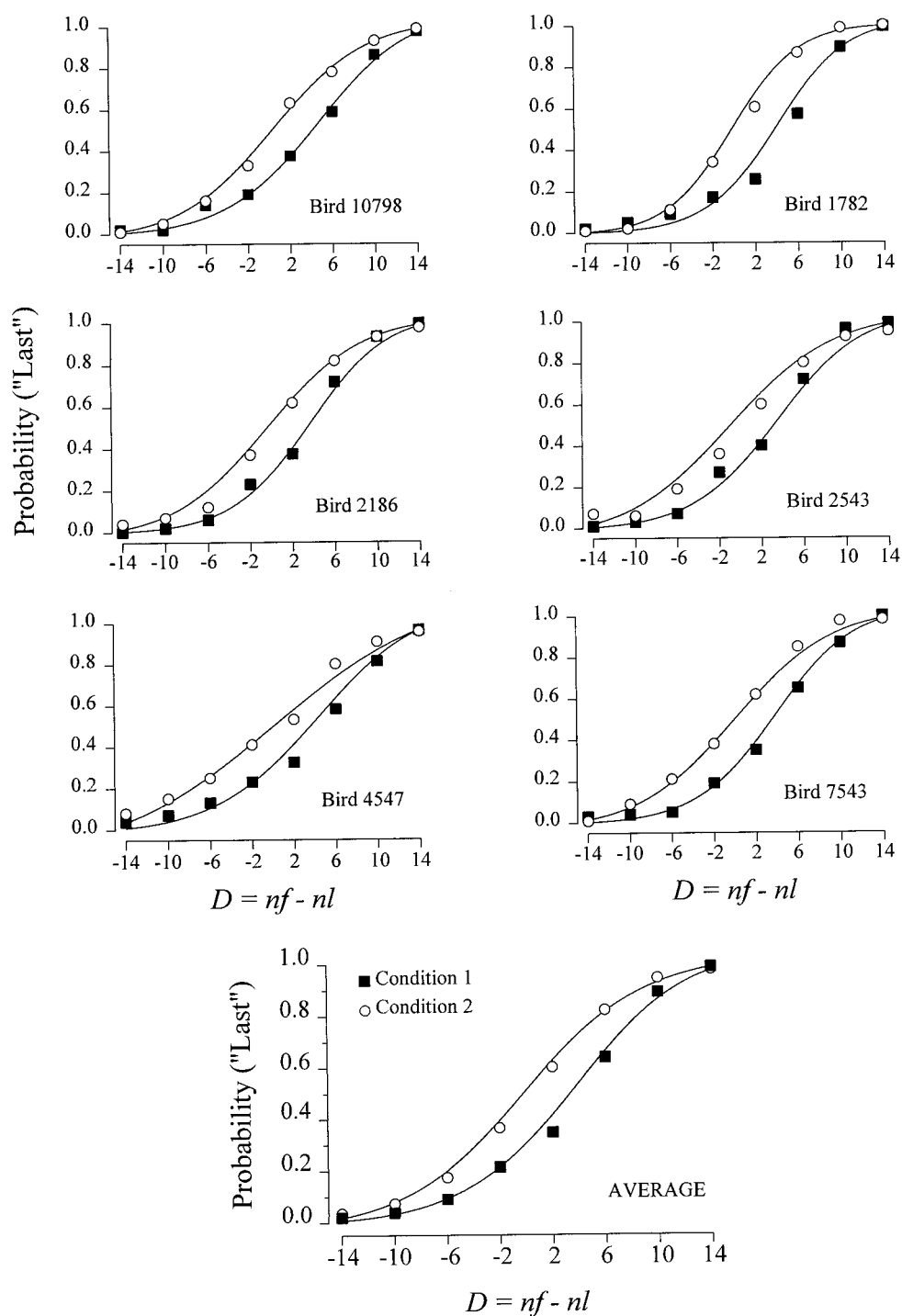


Fig. 7. The symbols show the probability of choosing the last stimulus of the sample as a function of the difference between the stimulus frequencies. The sample size was 16. The squares and circles are from Conditions 1 and 2, respectively. The top panels show the data from individual birds; the bottom panel shows the average across birds. The curves are the model's predictions using the parameters shown in Table 2.

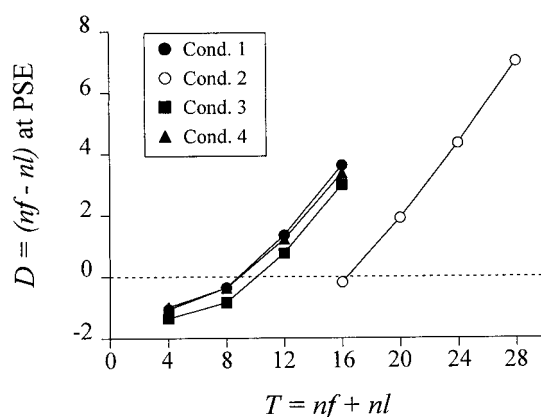


Fig. 8. Average value of D at the point of subjective equality [i.e., $p(\text{last}) = .5$] plotted against sample size, T .

probability of choosing the last stimulus decreased with T , and a repeated measures ANOVA yielded a statistically significant effect, $F(3, 15) = 11.6$, $p < .001$. For Birds 1782 and 2186, $p(\text{last})$ was significantly above .5 for the smallest sample; for Birds 1782, 4547, and 7543, $p(\text{last})$ was significantly below .5 for the largest sample (the 95% confidence interval associated with random responding is [.39 .61]).

The bottom left panel of Figure 9 shows the average results and the average curve predicted by the model. Again, data and curve suggest a primacy effect following the smallest sample and a recency effect following the largest one. The bottom right panel shows the estimated values of $p(\text{last})$ at $D = 0$ during Conditions 1, 2, and 3. In all cases, the curves decreased with the sample size, which means that the recency effect became more pronounced as the sample size increased. In addition, the curves from Conditions 1 and 3 were similar to the curve from Condition 4, and again they suggest a primacy effect after the smallest samples and a recency effect after the largest ones. Finally, the discontinuity between the curves from Condition 2 and the curves from Conditions 1 and 3 is another manifestation of the context effect.

In summary, the two ways of assessing primacy and recency effects provided convergent evidence: When the samples were small, the primacy effect tended to dominate; when the samples were large, the recency effect dominated. As the preceding figures might

have intimated, the model fitted the primacy and recency data well.

The Parameters of the Model

Figure 10 illustrates the behavior of the model's parameters across the experiment. The top left panel shows the values of the interference parameter for each pigeon. Although the magnitude of the changes was usually small, for 5 of the 6 birds (except 4547) γ decreased during Condition 2 and then increased during Condition 3 or 4. In other words, the degree of proactive interference varied directly with the average sample size. The top right panel shows the values of the decay parameter α . The changes in α across conditions were usually smaller in magnitude and less consistent in direction than the changes in γ .

The averages of γ and α for each experimental condition are shown in the bottom left panel. To avoid the overlap of the curves, the filled symbols were shifted upwards by 0.2. The filled circles reveal the decrease and subsequent increase in γ across conditions. A one-way repeated measures ANOVA yielded a significant effect of condition, $F(3, 5) = 4.32$, $p = .022$, and subsequent paired t tests showed that Conditions 1 and 4 differed significantly from Condition 2, $t(5) = 2.57$, $p < .05$, and $t(5) = 2.60$, $p < .05$. The open circles show that α did not change appreciably. A one-way repeated measures ANOVA yielded a nonsignificant effect of condition, $F(3, 5) = 2.09$, $p = 0.14$.

The bottom right panel plots the values of γ against the values of α in semilogarithmic coordinates. The lines are the best fitting regression lines. In all conditions, there was a significant negative correlation between the two parameters (the p values ranged from .0004 to .02). Hence, higher rates of decay were associated with stronger proactive interference.

A more extensive analysis revealed that the two model parameters were always negatively correlated, regardless of which conditions were compared (1 vs. 1, 1 vs. 2, etc., up to 4 vs. 4, for a total of 16 comparisons), whereas the correlations between the same parameter across different conditions (e.g., 1 vs. 2, 1 vs. 3, etc., for a total of six comparisons) always yielded positive correlations. The sign of these correlations, not their absolute value,

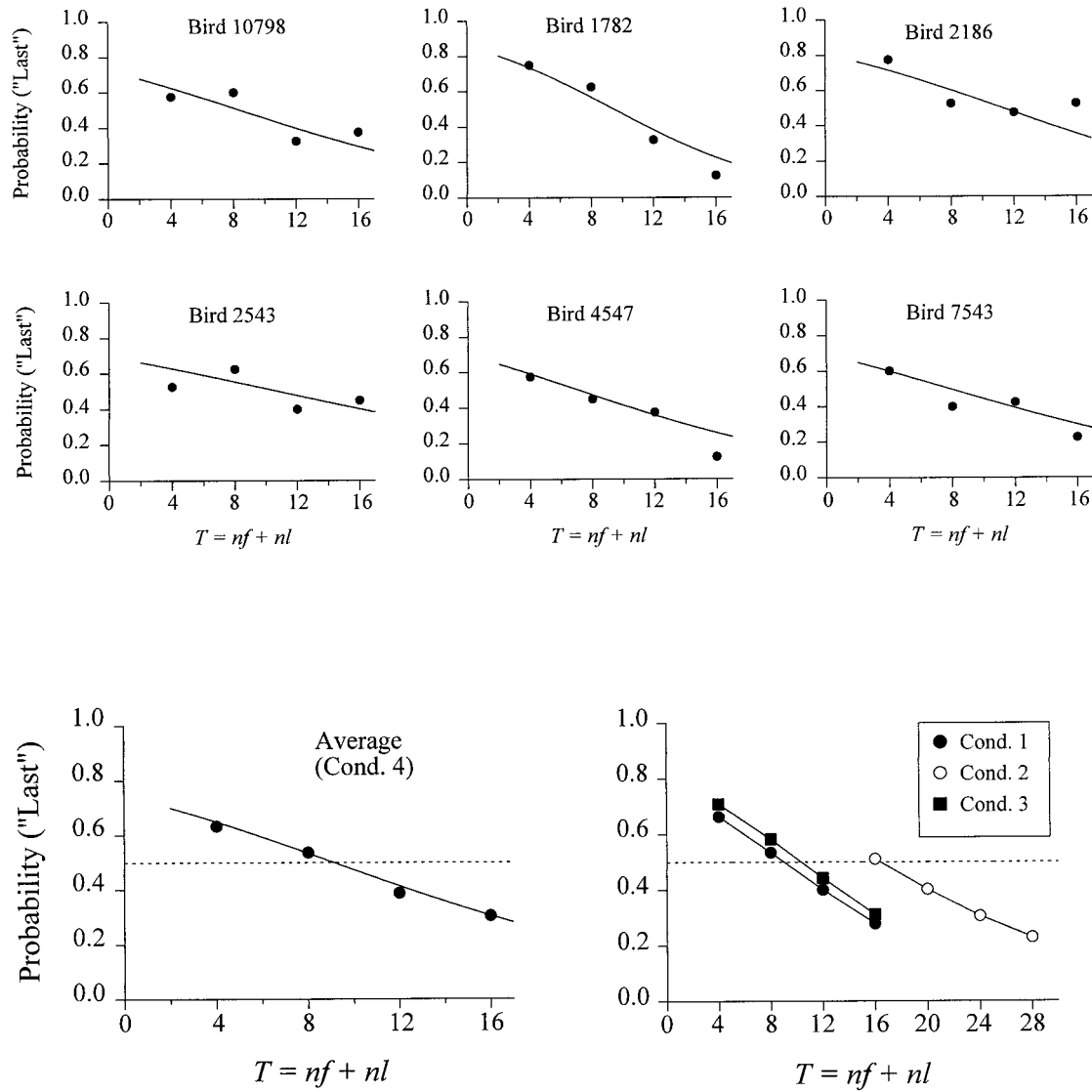


Fig. 9. Top: The symbols show the probability of choosing the last stimulus during Condition 4 as a function of the total sample size, T . The two stimuli had the same frequency (i.e., $D = 0$). The curves are the model's predictions with the parameters shown in Table 2. Bottom left panel: average data and fit for Condition 4. Bottom right panel: estimated $p(\text{last})$ values at $D = 0$ during Conditions 1, 2, and 3.

was completely consistent. We offer an interpretation of this finding below.

DISCUSSION

The purpose of this study was to understand how discrimination of relative frequency depends on two properties of the sample, the difference in the frequencies of the two stimuli, D , and the total number of stimuli in

the sample, T . To that effect, pigeons were presented with two series of lights, the first light with frequency nf the second with frequency nl . Choosing the least frequent stimulus of the sample was reinforced. Across four experimental conditions, D and T were varied systematically, independently, and over a relatively large range of values.

The results across the four conditions showed that pigeons learned the relative fre-

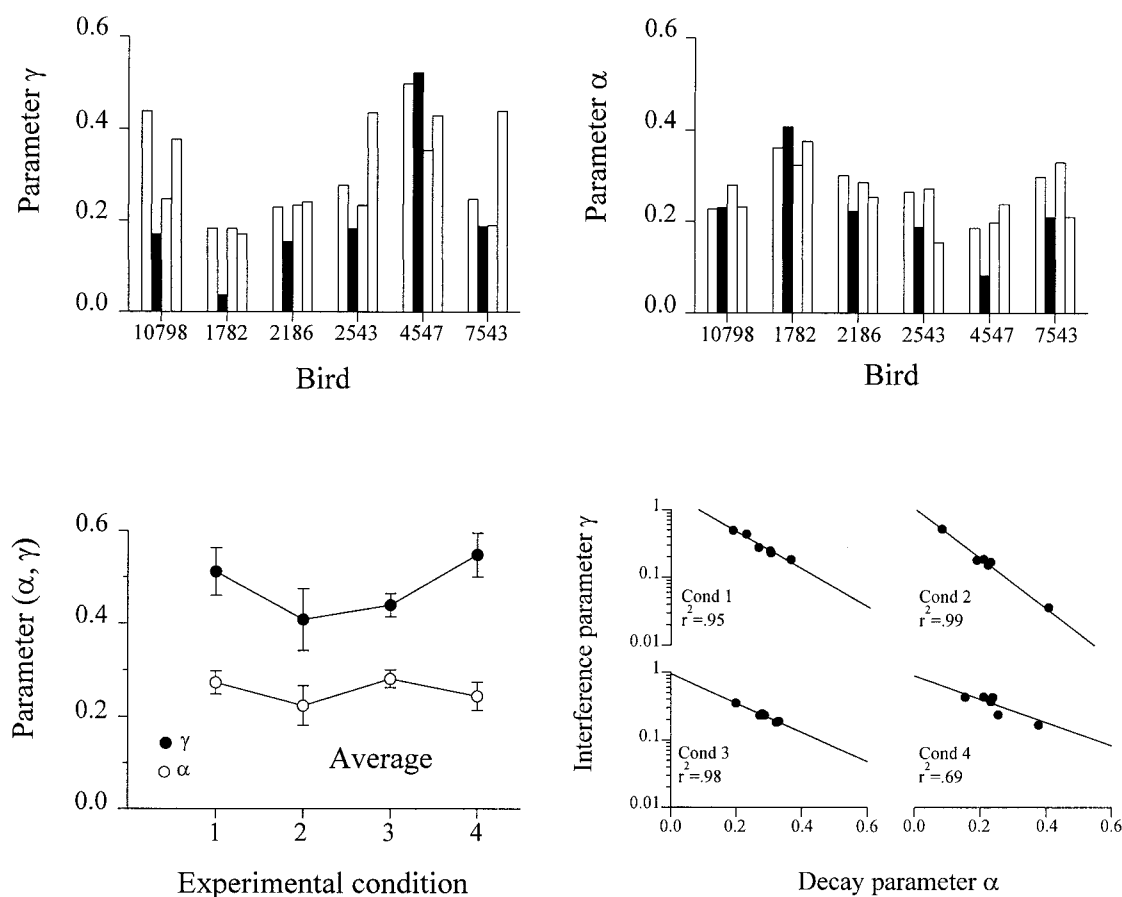


Fig. 10. Top: Value of the interference parameter γ (left) and decay parameter α (right) for each bird and condition. The bars are ordered from left to right according to experimental condition. Bottom left: The symbols show the average values of γ and α for each condition. The values for γ were shifted upwards by 0.2. Also shown are standard errors of the mean. Bottom right: regression of γ on α for each experimental condition. Each data point corresponds to 1 bird. The lines are best fitting regression lines. Note the logarithmic y axis.

quency discrimination well. Specifically, when $D > 0$, accuracy varied with the Weber-like fraction D/T : For a constant D , accuracy decreased with T , and, for a constant T , it increased with D . When $D < 0$, accuracy still improved with the difference between the stimulus frequencies, but the effect of the sample size was less clear. It was either too weak to be detected or, when it could be detected, it was the opposite of the effect observed for $D > 0$.

The discrimination of relative frequency also showed recency, primacy, and contextual effects. Small samples of four stimuli generally yielded a primacy effect in the sense that the first stimulus exerted greater control over the choice response than the last. However,

larger samples generally yielded a recency effect, in that the last stimulus exerted greater control over choice than the first one. In addition, performance on samples of 16 stimuli varied with the context in which these samples occurred. When the 16-stimuli samples were the largest samples of the set, performance showed a marked recency effect, but when they were the smallest samples of the set, the recency effect disappeared (see Figures 8 and 9). This result indicates that performance depends not only on the current trial but also on the trials surrounding it.

A simple quantitative model of the discrimination process was proposed. The model was based on the assumption that, properly speaking, the pigeon is not counting the stim-

ulus lights, performing mental arithmetic on numerons, remembering the number of lights at the moment of choice, or engaging in any similar process. Rather, we assumed that propensities to choose one or the other key simply changed as the sample stimuli are presented. In this sense, the model is prospective because it stresses the prospective controlling function of the sample stimuli on choice.

The model contained four key assumptions, all suggested by the work of previous researchers (e.g., Alsop & Honig, 1991; Grant & Roberts, 1973; Killeen, 1994; Machado & Cevik, 1997; Mechner, 1958; Meck & Church, 1983; Roberts & Grant, 1978; Wixted, 1990):

1. Choice probability depends on the ratio of the values of the control functions of the sample stimuli at the moment of choice.

2. The control function of a stimulus increases by a fixed amount with each presentation of the stimulus, and decays exponentially with each nonpresentation of the stimulus. The decay may be due to retroactive interference, the passage of time, or both. Its rate is measured by the parameter α .

3. The first stimulus may disrupt the control function of the second. The magnitude of such disruption, given by the parameter γ , measures the degree of proactive interference.

4. Because the task reinforces the choice of the least frequent stimulus, each stimulus exerts avoidance (as opposed to approach) control over the choice response. That is to say, if Stimulus A exerts more control than Stimulus B, then the bird is more likely to avoid A and choose B.

The model fit the data well. The variance accounted for was always above 96% (overall average was 99%), and the parameters of the model had reasonable and consistent values. The decay parameter α did not change appreciably across experimental conditions. Its overall average was 0.255, and because the sample stimuli occurred at a rate of approximately 1 stimulus per 0.8 s, the average time rate of decay equaled 0.32 per second.² The

interference parameter γ was always less than 1, which, according to the model, reveals the presence of proactive interference. Furthermore, there was some indication that the degree of proactive interference increased when the average sample size increased in Condition 2.

The decay and interference parameters were always negatively correlated in the population of pigeons. Thus, pigeons with small decay constants tended to also have lower degrees of proactive interference. This correlation makes sense, because if pigeons with smaller decay constants also showed larger degrees of proactive interference, then the second stimulus of the sample could hardly influence the choice response. The same functional argument may also explain why γ was never greater than 1 (i.e., another source of retroactive interference), because if that were to happen, the effects of interference and decay, instead of counteracting each other, would combine to disrupt the control function of the first stimulus. Good discrimination would be harder to achieve because the first stimulus of the sample would have little influence on the choice response.

If the assumption that α represents not only time decay but also retroactive interference is correct, then we have to conclude that the processes of retroactive and proactive interference followed different rules in the present experiment. Retroactive interference reduced the value of the control function of the first stimulus, S_F , in proportion to its current value. It caused, as it were, a global effect. In contrast, proactive interference reduced the effect of each occurrence of the second stimulus, a piecewise effect. This asymmetry may simply reflect the way the sample stimuli were presented (i.e., all red lights before or after the green lights). When the second stimulus started to occur, all instances of the first stimulus had been presented and S_F had been formed. Hence, each instance of the second stimulus could affect S_F in toto. For the same reason, S_F could change the effect of each instance of the second stimulus. Hence, we expect this asymmetry to vanish, or at least to be greatly attenuated, when the stimuli are intermixed in the sample.

The model also sheds light on the asym-

² The relation between α , the rate of decay per stimulus, and the time rate of decay per second, which we designate by λ , is $\lambda = \alpha/\delta$, where δ is the duration of the stimulus plus the interstimulus interval. λ has units S^{-1} , α is dimensionless, and δ has units S .

metry of the effects of T for $D > 0$ and $D < 0$. Recall that when $D > 0$ accuracy decreased with T , but when $D < 0$ accuracy either increased with T or was so close to 1 that no clear effect of T could be noticed. The first effect is a straightforward consequence of decay. Consider, for example, the two samples $nf = 11$, $nl = 5$ and $nf = 7$, $nl = 1$; for both, $D = +6$. In this case, the result of decay is to bring the effective value of nf closer to nl , which makes the discrimination harder. Moreover, because the amount of decay is proportional to nl , accuracy should be lower in the first sample. More generally, then, the model predicts that for $D > 0$, T should worsen performance, as it did.

Consider now the symmetric samples, $nf = 5$, $nl = 11$ and $nf = 1$, $nl = 7$; for both, $D = -6$. In this case, the result of decay is to take the effective value of nf away from nl , which should make the discrimination easier. Because the effect is more pronounced in the first sample (nl is greater), decay favors larger samples. However—and this is what makes matters more complicated when $D < 0$ —the ratio rule given by Equation 4 makes the difference of -6 harder to detect in the larger sample (cf. Equation 5 with $\alpha = 0$). Hence, we have two processes that oppose each other as the sample size increases. On the one hand, the choice rule makes the discrimination of the same difference harder; on the other hand, decay makes the same discrimination easier.

The offshoot of the preceding analysis is that when $D < 0$, the effect of T will depend on the absolute values of D , T , and α . In the Appendix we show that proportion correct will initially decrease with T but then it will increase. For large values of D , however, proportion correct will be so close to 1 that variations in T will have very small effects. In general, the data confirmed these predictions.

Frequency or cumulative duration? In the present experiment, the bird had to peck each sample stimulus to continue the trial and, consequently, the bird, not the experimenter, controlled the duration of the samples. This procedural arrangement may be criticized on the grounds that changes in sample duration may confound the interpretation of the results. As an alternative, one could have presented the sample stimuli independently of

the bird's behavior and therefore maintain sample duration constant. However, this alternative is not without its own disadvantages. For example, when stimuli are presented regardless of the animal's behavior, some stimuli may occur while the animal is not facing or looking at the keys. More generally, the observing response itself may become conditioned such that it will occur predominantly during the last segments of the sample. There is a methodological difficulty here. On the one hand, by advancing the trial only when the bird pecks the sample keys, the experimenter ensures that the animal sees each and every stimulus but loses control over the sample duration. On the other hand, by presenting the stimuli independent of behavior, the experimenter gains control over the stimulus duration but loses control over the observing response. It seems to us that the difficulty can be solved only by comparing the results obtained with both types of procedures against the predictions of quantitative models of the underlying process.

Nevertheless, some indirect evidence suggests that frequency rather than duration was the main controlling variable in the present experiment. We compared the results of multiple regression using either the cumulative durations of the two stimuli or the two stimulus frequencies as the independent variables. Cumulative duration was defined as the total time during which a keylight was on, or equivalently, the sum of all the latencies to peck the illuminated key. The results showed that, in most cases, frequency predicted performance better than cumulative duration. Specifically, in Conditions 1, 3, and 4, 17 of the 18 R^2 values ($6 \text{ birds} \times 3 \text{ conditions}$) were higher for stimulus frequency than for stimulus duration. The R^2 values could not be obtained for Condition 2 because a programming error during that condition caused the latencies after the 16th sample stimulus to be incorrectly saved to a file.

One should not conclude from the preceding remarks that temporal variables in general, and cumulative stimulus duration in particular, played no role in the frequency discrimination. For one thing, the conclusion would be illogical because the correlation between stimulus frequency and cumulative stimulus duration implies that if the former variable is related to accuracy in an

orderly way (as we know it was), then the latter must be also. Furthermore, the good fit of the model has already suggested the importance of temporal variables and their presumed effects (viz., the amount of decay of the control function of the first stimulus, S_F , during the occurrence of the second stimulus). Hence, one should conclude instead that both frequency and duration controlled performance, a conclusion that is also supported by other studies (e.g., Meck & Church, 1974; Roberts, 1995). It remains a challenge for a future theory to integrate frequency and duration explicitly into a real-time dynamic model of relative frequency discrimination.

The context effect. The model accommodated the context effect displayed in Figure 7 through parameter changes, and these changes suggested that the effect amounted to a greater influence on choice of the first stimulus of the sample. Obviously, this account is not a principled explanation—as a matter of fact, we had not anticipated the context effect—but is a post hoc interpretation of the data at the light of the model. Furthermore, the account leaves unanswered the next obvious question: Why did the parameters change during Condition 2? One possibility to account for either the parameter change or the context effect is that the duration of the 16-element samples may have differed between Conditions 1 and 2. To determine whether this was the case, we computed the median duration of the 16-element samples during the last 40 sessions of the two conditions. The results were inconsistent. Three pigeons experienced an appreciable increase in the sample duration from Condition 1 to Condition 2 (Bird 10798: from 14.6 s to 18.0 s; Bird 2543: from 9.7 s to 12.7 s; Bird 7543: from 12.2 s to 21.0 s), but the remaining 3 birds experienced no appreciable change (Bird 1782: from 12.8 s to 11.7 s; Bird 2186: from 10.0 s to 11.0 s; Bird 4547: from 11.9 s to 10.7 s). The overall medians for Conditions 1 and 2 were 12.05 s and 12.18 s, respectively. Although the changes in sample duration were inconsistent, all birds showed the context effect. We conclude that sample duration is not likely to explain either the context effect or the parameter change.

Another possibility³ is that during Condition 2 choices were based exclusively on the frequency of the last stimulus. If that frequency was below 8, they tended to choose the last stimulus; otherwise, they tended to choose the first stimulus. Of course, a similar rule based on the frequency of the first stimulus would work as well, but the point is that the discrimination may have depended on the absolute frequency of one of the stimuli, not on the relative frequencies of the two stimuli. In contrast, no absolute-frequency rule would guarantee success during Condition 1. Hence, a rule change from Condition 1 to Condition 2 may have brought about the context effect. Although plausible, this hypothesis has two shortcomings. First, even if true we would still not know how the hypothesis would explain differences in performance following the *same* $T = 16$ samples (see Figure 7). Second, and more important perhaps, some results seem to be inconsistent with the hypothesis. Take, for example, Samples 9-7, 13-7, 17-7, and 21-7 used during Condition 2. The hypothesis predicts the same proportion correct across the samples because the last stimulus has the same frequency. Instead, proportion correct usually increased with the difference between the two frequencies (see Figure 4). The results obtained with the mirror-image Samples 7-9, 7-13, 7-17, and 7-21 also seem to rule out a discrimination based exclusively on the frequency of the first stimulus.

A third hypothesis is that the control function of the first stimulus, S_F , decays not to 0 as the model assumed but to the average frequency of the first stimulus. Thus, in Condition 1 S_F would decay to 6 while the second stimulus occurred, whereas in Condition 2 it would decay to 10. To see how this regression towards the mean effect might account for the context effect, consider the samples shown in Table 3 and the corresponding averages of proportion correct.

Take Sample 7-9. In Condition 1, the frequency of the first stimulus would regress from 7 to 6, which would facilitate the discrimination because the difference between the effective frequencies at the moment of choice would increase from $7 - 9 = -2$ to $6 - 9 = -3$. In contrast, in Condition 2 the

³ We thank an anonymous reviewer for this suggestion.

Table 3

Average proportion (P) correct in Conditions 1 and 2 following some specific samples.

Sample	D	T	Condi- tion 1 P correct	Condi- tion 2 P correct
7-9	-2	16	.78	.61
9-7	+2	16	.35	.60
1-3	-2	4	.65	
3-1	+2	4	.84	

frequency of the first stimulus would regress from 7 to 10, which would make the discrimination harder. The regression hypothesis correctly predicts greater accuracy during Condition 1. Take now Sample 9-7. Regression toward 6 in Condition 1 would make the discrimination harder than the regression toward 10 in Condition 2. Again, the results are consistent with the prediction. A similar reasoning applied to all other samples of 16 stimuli would predict a decrease in accuracy when $D < 0$, but an increase when $D > 0$, during Condition 2. The results shown in Figure 7 support the prediction. Moreover, the hypothesis also predicts the primacy effect obtained with the smallest samples. Regression toward 6 would make the discrimination harder in Sample 1-3 than in Sample 3-1, which again agrees with the data.

Although the regression hypothesis predicts the context effect, it is less successful with other findings. Consider, for example, the pairs of samples 15-1/1-15, 11-1/1-11, and 7-1/1-7. According to the hypothesis, in Samples 1-15, 1-11, and 1-7 the frequency of the first stimulus would have ample opportunity to decay to 6, but this opportunity would be denied in Samples 15-1, 11-1, and 7-1. Hence, one would expect some difference in accuracy between the two samples of each pair. However, the accuracies were virtually identical (see bottom panel of Figure 3). More generally, without further assumptions (about the rate of regression, for instance), the hypothesis remains vague. Our attempts to specify quantitatively the regression model have been unsuccessful.

Extending the model: Alsop and Honig (1991). The present experiment was similar to Experiment 1 of Alsop and Honig's (1991) study and, therefore, we asked whether the

model could also account for their results. The authors presented pigeons with 7 flashes of blue or red light on a center key, and then reinforced choices of the right key when the red flashes outnumbered the blue and choices of the left key when the blue flashes outnumbered the red. Their average results are presented in the top panel of Figure 11. The curves show the proportion of "more blue" choices as a function of the proportion of blue flashes in the sample. For both types of samples, accuracy increased with the difference D between the stimulus frequencies. The relative position of the two curves reveals a strong recency effect, because the last stimulus exerted greater approach control over the choice response than the first.

Before we examine how the model can account for these findings, two other aspects of the data need to be discussed: the accuracy at the anchor points and the presence of bias. When the sample contained only blue or red stimuli (proportion of blue flashes equal to 0 or 1), accuracy was relatively low. Without some modification, the model cannot account for this result because it predicts perfect accuracy whenever one of the stimuli has zero frequency. A relatively poor discrimination following the presumably easiest samples may stem from a variety of reasons, such as the residual effect of previous trials, "guessing" whenever the bird happened to be distracted by extraneous stimuli, or even imperfect learning of which side key to choose following the different samples (see Blough, 1996; Heinman, Avin, Sullivan, & Chase, 1969). Although we cannot decide among these possibilities without further information, we illustrate how the model can be extended to deal with the last one. (The "guessing" alternative yielded virtually the same goodness of fit.)

Assume that on a proportion $1 - p$ of the trials the bird mistakenly reverses the key assignments. Hence, when the blue flashes precede the red ones, the probability of choosing the "more blue" key is given by

$$p(\text{more blue}) \mid \text{blue first} \\ = pz + (1 - p)(1 - z), \quad (7)$$

where z is the probability of more blue when no mistake occurs. The form of the term $(1 - p)(1 - z)$ stems from the fact that with

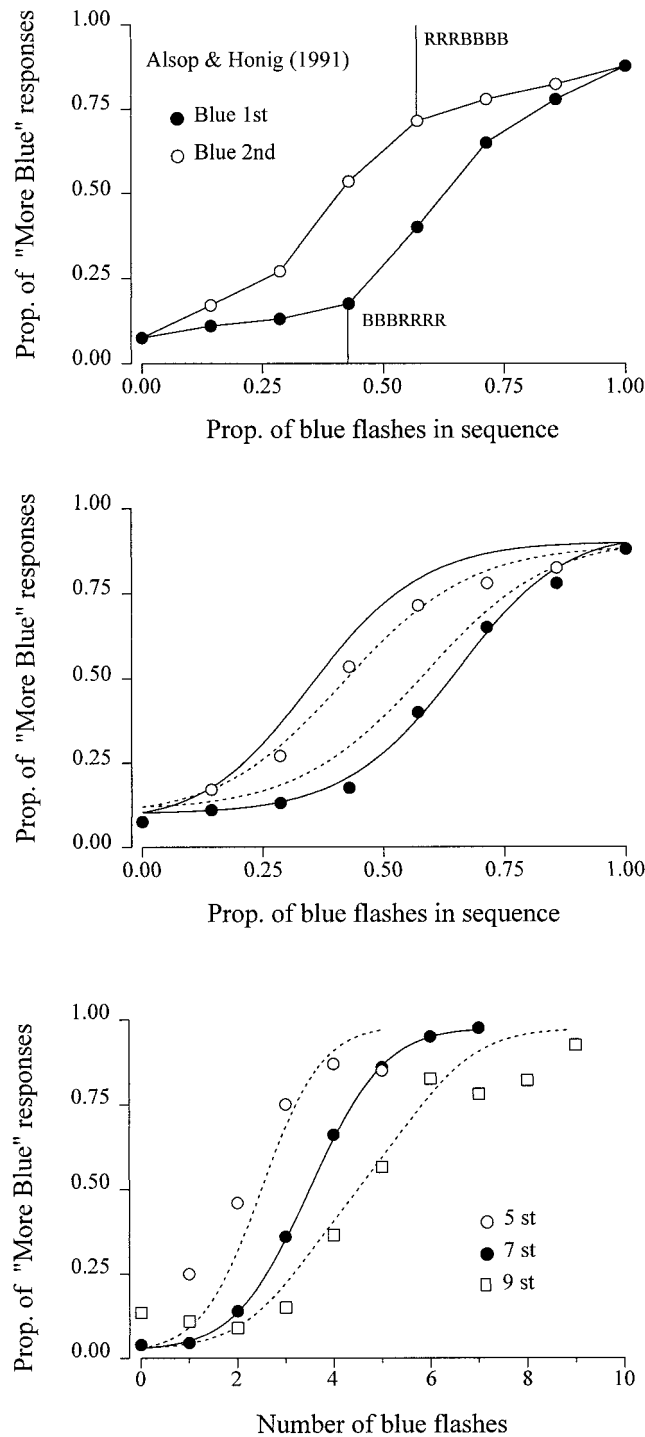


Fig. 11. Top: proportion of "more blue" responses in Alsop and Honig's (1991) study, Experiment 1. The filled circles represent samples in which blue was the first stimulus; the open circles represent the samples in which blue was the last. The vertical lines illustrate the presence of bias. Middle: model's predictions for the same data. The solid lines show the fit (Equations 7 and 8) to the blue first data; the dotted lines show the fit to the blue second data. Bottom: The solid curve shows the model's fit to the training samples of seven stimuli. The other curves show the predictions of the test results for samples with five and nine stimuli. The three curves used the same parameters: $\gamma = .09$, $\alpha = .86$, and $p = .97$.

probability $1 - z$ the choice of the “more red” key would have occurred, but because the bird confuses the key assignments, it actually pecks the “more blue” key. Given that blue is the first stimulus,

$$\begin{aligned} z &= \frac{S_F}{S_F + S_L} \\ &= \frac{\beta_1 \times nB \times \exp(-\alpha \times nR)}{\beta_1 \times nB \times \exp(-\alpha \times nR) + \beta_2 \times nR} \\ &= \frac{1}{1 + \gamma \frac{nR}{nB} \exp(\alpha \times nR)}, \end{aligned}$$

where nR and nB are the frequencies of the red and blue flashes, respectively. Following a similar argument one can show that when the red flashes precede the blue ones, the probability of choosing “more blue” is

$$\begin{aligned} p(\text{more blue} | \text{blue last}) \\ = p(1 - z') + (1 - p)z', \end{aligned} \quad (8)$$

where

$$z' = \frac{S_L}{S_F + S_L} = \frac{1}{1 + \gamma \frac{nB}{nR} \exp(\alpha \times nB)}.$$

The middle panel in Figure 11 replots the data with the model's predictions. Consider the solid lines. The line through the filled circles was obtained by fitting the data from the “blue first” samples, and it accounted for 99.4% of the variance. However, with the same parameter values the model overestimated the probability of choosing “more blue” when the blue flashes followed the red ones (see solid line above open circles). Conversely, the dotted line through the open circles was obtained by fitting the data from the “blue second” samples. Although the model accounted for 99.1% of the variance, the same parameters overestimated the data from the remaining samples (see dotted line above the filled circles).

The reason for the model's lack of fits stems from the presence of a systematic bias for the “more red” choice key. To see this, consider the two samples BBBRRRR and RRRBBBB. Without bias, the probability of “more blue” after the first sample should equal, or be close to, the probability of “more

red” after the second sample. But the vertical lines in the top panel of Figure 11 show that this equality did not hold—the probability of “more blue” after BBBRRRR (bottom vertical line) was smaller than the probability of “more red” after RRRBBBB (top vertical line). The same analysis for the remaining data points shows that the bias for the “more red” key was systematic. Because the model does not take bias into account, obviously it cannot fit the data from both types of samples simultaneously.

After additional training with samples of seven flashes, Alsop and Honig (1991) examined the birds' performances following samples of five or nine stimuli. At issue was whether the absolute or the relative number of blue flashes controlled the choice behavior. The bottom panel of Figure 11 shows their results, which clearly favored the latter hypothesis. The curves are the model's predictions. After fitting the model to the training data (filled circles; $\omega^2 = .999$), the same parameter values were used to predict the test results. Despite the discrepancies at the end points, the model reproduced the major trends in the data.

The results from this study showed that pigeons can learn to discriminate the relative frequencies of stimuli; that the discrimination is influenced by the difference between the frequencies of the sample stimuli and by the total number of stimuli in the sample; that the discrimination is modulated by recency and primacy effects, themselves a function of the sample size; and that the discrimination is also influenced by the overall context of a sample. A model that assumed that the controlling function of a stimulus increases linearly when the stimulus is presented, but decays exponentially when the stimulus is absent, accounted for the data well.

REFERENCES

- Alsop, B., & Honig, W. K. (1991). Sequential stimuli and relative numerosity discriminations in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 386–395.
- Blough, D. S. (1996). Error factors in pigeon discrimination and delayed matching. *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 118–131.
- Boysen, S. T., & Bernston, G. G. (1989). Numerical com-

- petence in a chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology*, 103, 23–31.
- Boysen, S. T., & Capaldi, E. J. (Eds.). (1993). The development of numerical competence: Animal and human models. Hillsdale, NJ: Erlbaum.
- Davis, H., & Perusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral and Brain Sciences*, 11, 561–615.
- Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior*. Boston: Allyn and Bacon.
- Gaffan, E. A. (1992). Primacy, recency, and the variability of data in studies of animals' working memory. *Animal Learning & Behavior*, 20, 240–252.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- Grant, D. S., & Roberts, W. A. (1973). Trace interaction in pigeon short-term memory. *Journal of Experimental Psychology*, 101, 21–29.
- Heinman, E. G., Avin, E., Sullivan, M. A., & Chase, S. (1969). Analysis of stimulus generalization with a psychophysical method. *Journal of Experimental Psychology*, 80, 215–224.
- Hitch, G. J. (1983). Short-term memory processes in humans and animals. In A. Mayes (Ed.), *Memory in animals and humans: Some comparisons and their theoretical implications* (pp. 107–122). Workingham, UK: Van Nostrand Reinhold.
- Honig, W. K., & Matheson, W. R. (1995). Discrimination of relative numerosity and stimulus mixture by pigeons with comparable tasks. *Journal of Experimental Behavior: Animal Behavior Processes*, 21, 348–363.
- Killeen, P. R. (1994). Mathematical principles of reinforcement. *Behavioral and Brain Sciences*, 17, 105–172.
- Machado, A., & Cevik, M. (1997). The discrimination of relative frequency by pigeons. *Journal of the Experimental Analysis of Behavior*, 67, 11–42.
- Mechner, F. M. (1958). Probability relations within response sequences under ratio reinforcement. *Journal of the Experimental Analysis of Behavior*, 1, 109–121.
- Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 320–334.
- Pepperberg, I. M. (1987). Evidence for conceptual quantitative abilities in the African grey parrot: Labeling of cardinal sets. *Ethology*, 75, 37–61.
- Roberts, W. A. (1995). Simultaneous numerical and temporal processing in the pigeon. *Current Directions in Psychological Science*, 4, 47–51.
- Roberts, W. A., & Grant, D. S. (1974). Short-term memory in the pigeons with presentation time precisely controlled. *Learning and Motivation*, 5, 393–408.
- Roberts, W. A., & Grant, D. S. (1978). Interaction of sample and comparison stimuli in delayed matching to sample with pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 4, 68–82.
- Wixted, J. T. (1990). Analyzing the empirical course of forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 927–935.
- Wright, A. A. (1994). Primacy effects in animal memory and human nonverbal memory. *Animal Learning & Behavior*, 22, 219–223.
- Wright, A. A., Santiago, H. C., Sands, C. F., Kendrick, D. F., & Cook, R. G. (1985). Memory processing of se-

rial lists by pigeons, monkeys, and people. *Science*, 229, 287–289.

Received December 9, 1998
Final acceptance April 20, 1999

APPENDIX

Retroactive interference plus time decay. Assume that when the second stimulus occurs, $S_F = S$. The first instance of the second stimulus reduces S_F instantaneously, as it were, to a fraction of its value, say ρS_0 , with $0 \leq \rho \leq 1$. This reduced value then decays at some rate κ until the next instance of the stimulus occurs. After nl instances of the second stimulus,

$$S_F = (S \times \rho^{nl}) \times \exp(-\kappa \times t \times nl),$$

where t is the duration of each occurrence of the second stimulus. The first term in parentheses shows the effect of retroactive interference, and the second shows the effect of decay.

But the preceding expression is of the form

$$S_F = S \times \exp(-\alpha \times nl)$$

for some value of α . We obtain Equation 2 in the text.

Partial derivatives of $p(\text{last})$. To see how $p(\text{last})$ changes with D and T , we rewrite Equation 5 in terms of $D = nf - nl$ and $T = nf + nl$ and then take partial derivatives. Letting p stand for $p(\text{last})$, we get

$$\frac{\partial p}{\partial D} = p(1 - p) \left[\frac{2T}{T^2 - D^2} + \frac{\alpha}{2} \right].$$

Because $D < T$, the term in brackets is always positive and therefore $\partial p / \partial D > 0$. For $D > 0$, proportion correct equals p , and therefore proportion correct increases with D . For $D < 0$, proportion correct equals $1 - p$, and therefore, as D increases in absolute value, proportion correct also increases. In short, proportion correct increases with the absolute value of D .

For T we get

$$\frac{\partial p}{\partial T} = -\gamma p(1 - p) \left[\frac{2D}{T^2 - D^2} + \frac{\alpha}{2} \right].$$

For $D > 0$, the term in brackets is positive and therefore $\partial p / \partial T < 0$; hence, proportion

correct decreases with T . However, when $D < 0$ the sign of $\partial p / \partial T$ will depend on the absolute values of D , T , and α . Consider two extreme cases, $\alpha = 0$ (no decay) and $\alpha = 1$ (rapid decay; see Table 2 for obtained values of α). In the former case, $\partial p / \partial T > 0$ and proportion correct, equal to $1 - p$, decreases with T . In the latter case, $\partial p / \partial T < 0$

$$\left(\text{because } -\frac{1}{2} < \frac{2D}{T^2 - D^2} = \frac{nf - nl}{2 \times nf \times nl} < 0 \right),$$

and proportion correct increases with T . For intermediate cases ($0 < \alpha < 1$), $\partial p / \partial T$ changes sign from positive to negative as T increases. Hence, proportion correct initially decreases with T but then increases.